Published in Australia, New Zealand, the Pacific, Asia and Africa by
Longman Cheshire Pty Limited
Longman House
Kings Gardens
95 Coventry Street
Melbourne 3205 Australia
ISBN 0 582 71286 6

Offices in Sydney, Brisbane, Adelaide and Perth. Associated companies, branches and representatives throughout the world.

Copublished in the Western Hemisphere, United Kingdom and Europe by Halsted Press: an Imprint of
John Wiley & Sons, Inc.
New York Toronto Chichester

First published 1992

Designed by Nadia Graziotto
Printed in Hong Kong

National Library of Australia
Cataloguing-in-Publication data

Time-frequency signal analysis.
Bibliography.
Includes index.
ISBN 0-582-71286-6
1.Signal processing. I. Boashash, Boualem.
621. 38223

Library of Congress Cataloguing-in-Publication data

Time-frequency signal analysis-methods and applications/edited by Boualem Boashash.
p. cn.
Includes bibliographical references and index.
ISBN 0-470-21821-5
1. Signal processing. I. Boashash, Boualem.
YK5102.T585 1992
621.382'2-dc20 91-32658
CIP

TIME - FREQUENCY SIGNAL ANALYSIS

METHODS AND APPLICATIONS

Edited by Boualem Boashash

Longman Cheshire

WILEY HALSTED PRESS

Process., 38, no. 7, pp. 1084-1091, 1990.

[19] S. Oh and R. J. Marks II, "Some Properties of the Generalized Time-Frequency Representation With Cone Shaped Kernel", preprint 1990.

[20] S. Oh, R. J. Marks II, L. E. Atlas and J. A. Pitton, "Kernel Synthesis for Generalized Time-Frequency Distributions Using the Method of Projection Onto Convex Sets", *Advanced Algorithms and Architectures for Signal Processing III*, Franklin T. Luk, Editor, Proc. SPIE 1990, to appear.

[21] R. G. Baraniuk and D. L. Jones, "Optimal Kernels for Time-Frequency Analysis", *Advanced Algorithms and Architectures for Signal Processing III*, Franklin T. Luk, editor, Proc. SPIE 1990, to appear.

[22] E. F. Velez and R. G. Absher, "Transient analysis of Speech Signal Using the Wigner Time-Frequency Representation", ICASSP-89, pp. 2242-2245, 1989.

[23] E. F. Velez and H. Garudadri, "Speech Analysis Based on Smoothed Wigner-Ville Distribution", this volume.

[24] H. Garudadri, M. P. Beddoes, A. P. Benguerel and J. H. V. Gilbert, "On Computing the Smoothed Wigner Distribution", *Proc. IEEE Int. Conf. Acoust., Speech, Signal. Process.*, 87 pp. 1521-1524, 1987.

[25] W. Martin and P. Flandrin "Wigner-Ville Spectral Analysis of Nonstationary Processes", *IEEE Trans. Acoust. Speech Signal Process.*, pp. 1461-1470, 1985.

[26] L. B. White, "Some Aspects of the Time-Frequency Analysis of Random Processes", Thesis, University of Queensland, 1989.

[27] M. G. Amin, "Time-Frequency Spectrum Analysis and Estimation For Non-Stationary Random Processes", this volume and references therein.

[28] J. Bertrand and P. Bertrand, "Affine Time-Frequency Distribution", this volume.

Chapter 2

Instantaneous Frequency and Time-Frequency Distributions

Boualem Boashash and Graeme Jones

Keywords: instantaneous frequency, group delay, bandwidth-duration product, analytic signal, Hilbert transform, quadrature signal, digital instantaneous frequency

1 Introduction

This chapter presents the fundamental principles on which the notions of instantaneous frequency (IF), group delay (GD) and analytic signal are based, in the light of their relationship with time-frequency distributions (TFDs). An understanding of these notions is necessary in order to efficiently use them to derive methods for time-frequency analysis of non-stationary signals.

As indicated in chapter 1, many natural and man-made signals have spectral characteristics which vary in time. The chirp (linear FM) signal is such an example: it consists of a sine wave whose frequency increases (or decreases) linearly with time. Such a signal is used in practical situations such as seismic surveying, communications, radar, sonar. It also occurs naturally in such places as the echolocation systems of bats.

For the chirp and other time-varying signals the instantaneous frequency (IF) is an important descriptor. Qualitatively, this is a parameter which corresponds to the frequency of a sine wave which locally (in time) fits the signal under analysis. It has meaning only for monocomponent signals, that is, where there is only one frequency or a narrow range of frequencies varying as a function of time. In other

cases, the notion of a single valued IF becomes meaningless, and some form of appropriate break-down into sub-components is necessary.

2 Basic concepts

2.1 Evolution of the term 'instantaneous frequency'

This section presents a review of some early work concerned with defining and interpreting the IF. Carson and Fry in 1937 [1] generalised the notion of frequency by allowing it to vary as a function of time. They argued that it should be termed the IF, since it was a generalisation of the definition for constant frequency i.e. the rate of change of the phase angle at time t.

In 1946 Van der Pol [2], in the context of communication theory, defined the IF by considering a signal of the following form:

$$s(t) = a_0 \cos(2\pi f t + \phi_0) \tag{1}$$

where a_0 is a constant, f is the frequency of the oscillation, and ϕ_0 is a phase constant.

In this expression, Van der Pol makes a_0 and ϕ_0 vary and defines the amplitude and phase modulations as:

$$a(t) = a_0[1 + kh(t)] \tag{2}$$

and

$$\phi(t) = \phi_0[1 + mg(t)] \tag{3}$$

where k and m are constants, and $h(t)$ and $m(t)$ are the time-varying modulating waveforms. One could then express the frequency as:

$$f = f_0[1 + mg(t)] \tag{4}$$

Equation (4), however, leads to physical inconsistences as there is time-varying information in the modulated phase. This may be seen by the substitution of f in (4) back into equation (1).

Van der Pol argued that to properly obtain an expression for the IF of a phase modulated signal, equation (1) should be re-written as:

$$s(t) = a_0 \cos\left[\int_0^t 2\pi f_i(t)dt + \phi_0\right] \tag{5}$$

To further explain the concept, Van der Pol used an analogy with harmonic motion. For a vector rotating at constant angular velocity the frequency is:

$$f = \frac{1}{2\pi}\frac{d\phi(t)}{dt} \tag{6}$$

where $\phi(t) = 2\pi f t$ is the phase, and may be considered to be the angle the rotating vector forms with the horizontal reference line. He argued that for a varying angular frequency (represented by a vector rotating with a variable angular velocity), one could still use the definition given in (6) and obtain an (instantaneous) frequency of rotation:

$$f_i(t) = \frac{1}{2\pi}\frac{d\phi(t)}{dt} \tag{7}$$

This vector may be visualized as the complex signal (see figure 2.1):

$$z(t) = a_0 e^{j\left[\int_0^t 2\pi f_i(t)dt + \phi_0\right]} \tag{8}$$

which is the sum of the real signal in (5) and the pure imaginary quadrature signal:

$$j\sigma(t) = ja_0 \sin\left[\int_0^t 2\pi f_i(t)dt + \phi_0\right] \tag{9}$$

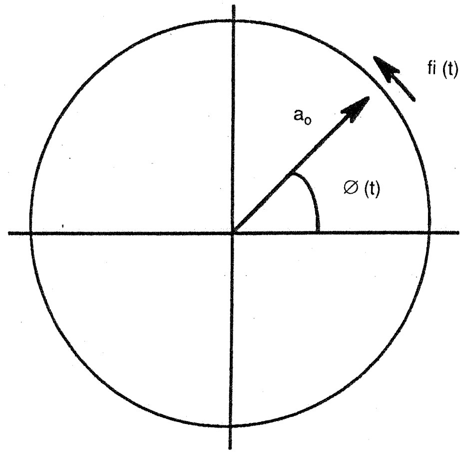

Figure 2.1. A vector (complex signal) rotating with variable angular velocity (instantaneous frequency).

Thus the phase of the resultant complex signal is used to define the IF of (7) which is consistent with the generally accepted definition.

Gabor [3] proposed the Hilbert transform as a means of obtaining a unique complex signal from a real one — with the imaginary part in quadrature to the real part (equation 9). The Hilbert transform is defined by:

$$\mathcal{H}[s(t)] = \text{p.v.}\left[\int_{-\infty}^{+\infty} \frac{s(t-\tau)}{\pi\tau}d\tau\right] \tag{10}$$

where p.v. denotes the Cauchy principal value of the integral. Ville [4] used this transform to generate his complex analytic signal. With the IF as defined in (7), he showed that the average frequency (using Gabor's average measures [3]) in a signal's spectrum was equal to the time average of the IF [4]:

$$< f >=< f_i > \tag{11}$$

where

$$< f >= \frac{\int_{-\infty}^{+\infty} f|Z(f)|^2 df}{\int_{-\infty}^{+\infty} |Z(f)|^2 df} \tag{12}$$

and

$$< f_i >= \frac{\int_{-\infty}^{+\infty} f_i(t)|z(t)|^2 dt}{\int_{-\infty}^{+\infty} |z(t)|^2 dt} \tag{13}$$

The analytic signal is $z(t) = s(t) + \mathcal{H}[s(t)]$, and $Z(f)$ is the Fourier transform of $z(t)$.

The interpretation of the IF has been a controversial issue. For example, Shekel [6] argued that the IF defined by:

$$f_i(t) = \frac{1}{2\pi}\frac{d\phi(t)}{dt} \tag{14}$$

where $z(t) = a(t)e^{j\phi(t)}$ is not a unique function of time, since, for example, any AM wave (written in complex form) may be expressed as either $m(t)e^{j2\pi ft}$ or $m_0 e^{j\phi(t)}$, where m_0 is a constant. A unique complex representation is obtained by using the HT, as Gabor and Ville have stated [3] [4], but whether or not it corresponds to any physical reality is another question, which is discussed in a later section.

Mandel [7] also challenged any physical interpretation of the IF by providing examples where the IF of a signal does not correspond to any of its Fourier spectral components. In addition, he showed that for moments higher than the first, the Fourier frequency and the IF average measures do not coincide, as Ville had noted for the second moment [4]. Several other authors have also contributed to the study of IF in more recent work [8] [9] [10] [11].

2.2 The Hilbert transform and the analytic signal

Gabor first used the Hilbert transform (HT) to generate his complex signal according to:

$$z(t) = s(t) + jy(t) \tag{15}$$

where $y(t)$ represents the HT of $s(t)$, as defined in (10). This signal is referred to as the analytic signal. It does not always correspond, however, to a signal plus its imaginary quadrature component. As an illustration, suppose we have a signal of the form:

$$s(t) = \Pi_T(t)\cos(2\pi f_0 t + \phi_0) \tag{16}$$

where $\Pi_T(t)$ denotes a box function of unit amplitude, duration T and symmetric about $t = 0$.

By inspection, it is obvious that the signal in phase quadrature is:

$$q(t) = \Pi_T(t)\sin(2\pi f_0 t + \phi_0) \tag{17}$$

Applying the HT to the signal of (16) does not exactly yield the signal in phase quadrature (equation 17). This is because the HT operation preserves the positive frequency domain of the spectrum and inverts the sign of the spectrum in the negative frequency domain. Thus in this case it does not simply transform the cosine term into a sine term.

It is therefore essential to know when the HT generates the signal's quadrature component exactly. This can be determined by considering a signal of the form $a(t)\cos\phi(t)$, and defining conditions for the following equation to be verified:

$$\begin{aligned} a(t)\cos\phi(t) + j\mathcal{H}[a(t)\cos\phi(t)] &= a(t)e^{j\phi(t)} \\ &= a(t)(\cos\phi(t) + j\sin\phi(t)) \end{aligned} \tag{18}$$

Equation (18) is satisfied if one of the following two conditions holds:

1 The spectrum of $a(t)$, $A(f)$, lies entirely in the region $|f| < f_0$ and $\mathcal{F}\{\cos\phi(t)\}$ only exists out of this region [12] [13]; and

2 $A(f)$ is non-zero only for $f > -f_1$, and $\mathcal{F}\{\cos\phi(t)\}$ is non-zero only for $f > f_2$; for $f_2 \geq f_1 \geq 0$ [4] [14] [35] where $\mathcal{F}$ denotes the Fourier transform operation.

These two conditions are obtained by examining the result of the convolution of $A(f)$ and $\mathcal{F}\{\cos\phi(t)\}$, as $\mathcal{F}\{a(t)\cos\phi(t)\} = A(f) * \mathcal{F}\{\cos\phi(t)\}$. It should also be stated that condition 2 above is not a practical constraint as $a(t)$ is defined to be an amplitude envelope and thus it must be low frequency, centred at the DC point.

The HT-generated analytic signal is meaningful and representative of the physical signal only under some conditions. As explained above,

if we have a signal with an imposed modulation of the form $a(t)\cos\phi(t)$ where physical meaning is attached to $a(t)$ and $\phi(t)$, and if the spectra of $a(t)$ and $\cos\phi(t)$ are not separated in frequency, then the HT will not be able to distinguish the amplitude and phase functions. Although the analytic signal will be of the form:

$$a_z(t)e^{j\phi_z(t)} \tag{19}$$

and will be unique, $a_z(t)$ and $\phi_z(t)$ may have no practical meaning. The more closely a signal approaches a narrowband condition, the better is the analytic signal approximation to the quadrature signal.

2.3 The bandwidth-duration(BT) product

The previous work examining the HT, analytic signal and the IF has not considered practical aspects of signals. In practice, signals have imposed constraints such as causality, finite energy, finite duration and stability. Often, it is also assumed that they possess a finite spectral bandwidth (within which most of the signal energy is contained). Although a signal with both a finite duration T and a finite bandwidth B is prohibited by the Fourier transform, signals which approach this double limitation form an important class. A review of such concepts and measures for time-frequency bounds are given in [15].

When signals have a BT product sufficiently high, the approximation error involved in assuming band and time limited functions is very small. These signals are often referred to as 'asymptotic' [16], the 'asymptotic' behaviour being measured by the parameter BT.

With these notions in mind, the results involving the HT and the analytic signal from the previous section may now be put in a more general and quantitative framework, which is summarized in the following theorems:

Theorem 1
If the signal, $s(t) = a(t)\cos\phi(t)$, has a monotonic and positive frequency law, then the quantity $s(t) + j\mathcal{H}[s(t)]$ approaches $z(t) = a(t)e^{j\phi(t)}$ asymptotically as BT $\rightarrow \infty$.

The proof follows work by Bedrosian [12] and Vakman [17] and is based on the following relation holding when BT $\rightarrow \infty$.

$$\mathcal{H}[s(t)] = \mathcal{H}[a(t)\cos\phi(t)] = a(t)\mathcal{H}[\cos\phi(t)] = a(t)\sin\phi(t) \tag{20}$$

It should be noted that the theorem states that the IF must be positive for the duration of the signal, and it must be monotonic in time [18].

Theorem 2
For a signal of the form, $s(t) = a(t)\cos(2\pi f_0 t + \phi(t))$ (f_0 is the constant or tonal frequency, not present in $\phi(t)$), with an arbitrary BT value, the analytic signal derived using a Hilbert transform asymptotically approaches the signal plus its imaginary quadrature part, i.e., $a(t)e^{j(2\pi f_0 t+\phi(t))}$, as $f_0 \rightarrow \infty$.

The proof follows work by Nuttall [14], where he shows that the difference in energy between the HT generated analytic signal and the signal plus its imaginary quadrature part is the energy in the spectrum of:

$$S_0(f) = \mathcal{F}\{a(t)e^{j\phi(t)}\} \tag{21}$$

for $f < -f_0$. Thus the representation becomes exact as $f_0 \rightarrow \infty$.

Theorem 2 indicates that the approximation error in forming the complex signal with the HT (compared to the signal plus its imaginary quadrature component) and the BT product are not directly related. In effect, theorem 1 is a special case of the more general theorem 2. A good approximation to the $a(t)e^{j(2\pi f_0 t+\phi(t))}$ term (and thus the IF) can be achieved even for a low BT product signal, provided its centre frequency is large. Two examples are used here to illustrate the theorems.

Example 1: Consider two signals of the form,

$$s(t) = \Pi_T(t)\cos 2\pi\left(f_0 t + \frac{\alpha}{2}t^2\right) \tag{22}$$

Signal 1: $f_0 = 50$Hz, $\alpha = 1714.3$, BT $= 2.1$
Signal 2: $f_0 = 50$Hz, $\alpha = 93.7$, BT $= 38.1$

Figure 2.2 shows the (analytic) spectrum of signal 1 (created using the HT). This may be compared to figure 2.3 which is the spectrum of signal 1 plus its phase quadrature component. Figures 2.4 and 2.5 show the same results for signal 2. Examination of the figures provides a simple illustration of theorem 1 — the analytic signal approaches the real plus quadrature signal as the BT becomes large (this may be seen for the spectra of signal 2).

Example 2: Consider two signals, once again, of the general form of (22):

Signal 1: $f_0 = 11$Hz, $\alpha = 333.3$, BT $= 1.2$
Signal 2: $f_0 = 51$Hz, $\alpha = 333.3$, BT $= 1.2$

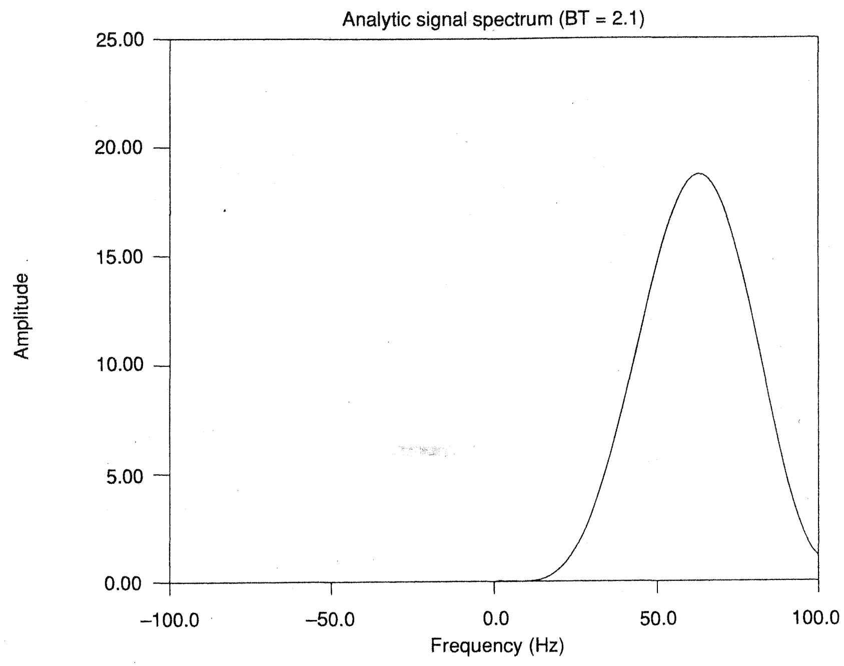

Figure 2.2. The analytic signal spectrum for signal 1 of example 1.

In a similar manner to the first example, figures 2.6(2.8) and 2.7(2.9) give the analytic spectrum and the real plus quadrature spectrum respectively for signal 1 (2). From the figures it is obvious that the analytic signal approaches the real plus quadrature complex signal as f_0 becomes larger, which illustrates theorem 2.

2.4 Instantaneous frequency and group-delay

Let $s(t)$ be a continuous time real signal with $z(t)$ its corresponding analytic signal:

$$z(t) = a(t)e^{j\phi(t)} \tag{23}$$

The IF is simply obtained by differentiation of the phase and scaling by $\frac{1}{2\pi}$. $z(t)$ has a complex spectrum which may be expressed as:

$$Z(f) = A(f)e^{j\theta(f)} \tag{24}$$

where $a(t)$ and $A(f)$ are positive functions. Another quantity of interest for the signal is the group delay (GD) defined by:

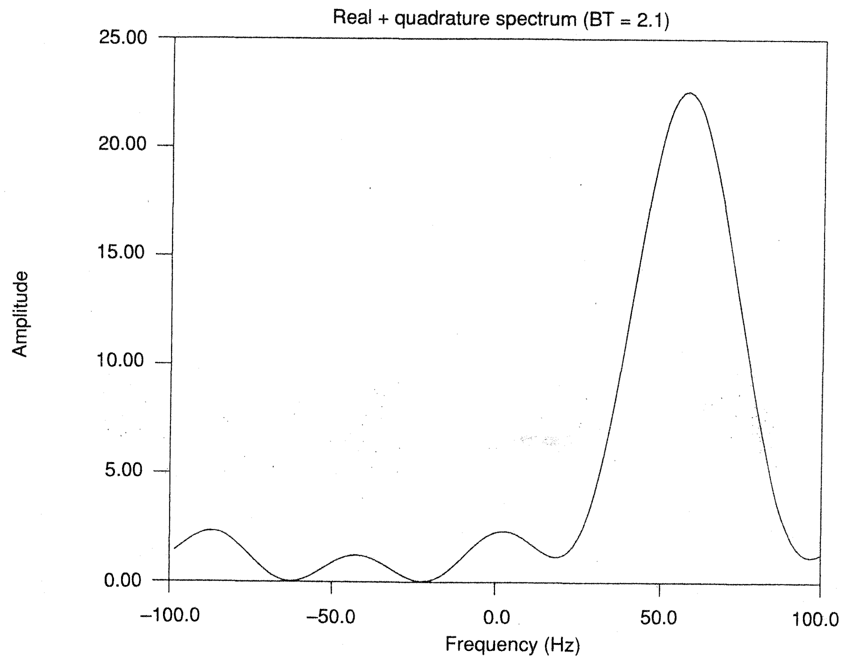

Figure 2.3. The real signal plus quadrature signal spectrum for signal 1 of example 1.

$$\tau_g(f) = -\frac{1}{2\pi}\frac{d\theta(f)}{df} \tag{25}$$

In many applications it is common to use the GD, $\tau_g(f)$, of the signal, to characterise the time-frequency law of the signal. It is therefore desirable to relate the two quantities of IF and GD.

In order to achieve this, the Fourier transform of the signal is examined. For signals of the form, $a(t)e^{j\phi(t)}$, with a large BT product and a monotonic IF law, the Fourier transform may be approximated by application of the stationary phase principle as follows [17]:

$$\mathcal{F}\{a(t)e^{j\phi(t)}\} \approx \left[\frac{2\pi}{|\phi''(\lambda)|}\right]^{\frac{1}{2}} a(\lambda)e^{j[-2\pi f\lambda+\phi(\lambda)\pm\pi/4]} \tag{26}$$

where λ is the stationary phase point, that is, the solution of :

$$\frac{d}{dt}(-j2\pi ft + \phi(t)) = 0 \tag{27}$$

It is simple to show, using equation (26), that the IF and GD are approximately inverses of each other (i.e., $f_i(t) = \tau_g^{-1}(f)$) [18] for these

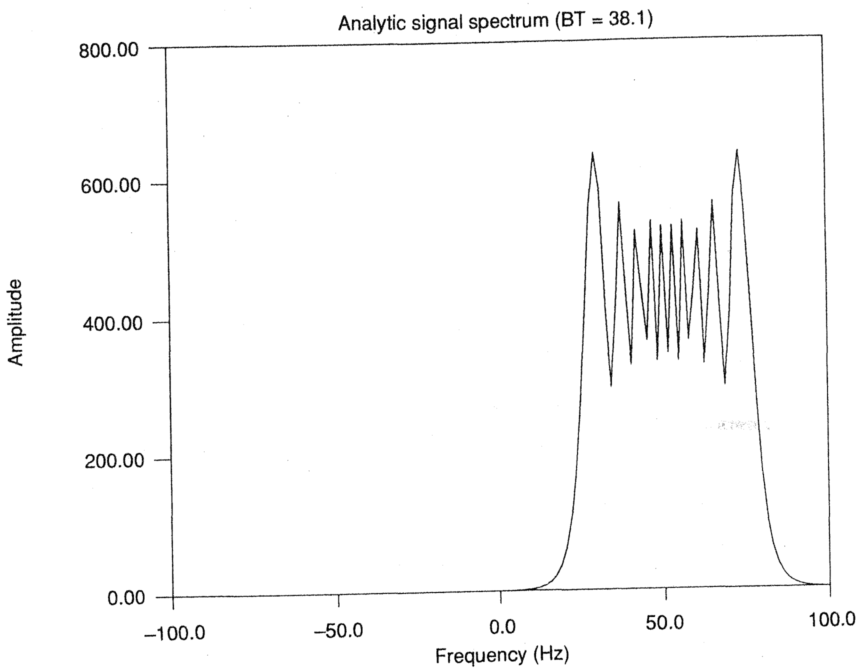

Figure 2.4. The analytic signal spectrum for signal 2 of example 1.

large BT signals. A further discussion of the relationship between IF and GD (including analytical examples) may be found in [15].

Signal concentration and elementary cells

A related concept utilizing IF and GD laws is that of "elementary cells" which define the signal concentration about these laws. One way to define the dimensions of these cells is to introduce the notion of "relaxation time", T_r, and "dynamic bandwidth," B_d, as [8]:

$$T_r(t) = \left| \frac{df_i(t)}{dt} \right|^{1/2} \tag{28}$$

$$B_d(f) = \left| \frac{d\tau_g(f)}{df} \right|^{1/2} \tag{29}$$

These measures represent boundaries in time and frequency within which most of the signal energy is contained. They provide quantitative support to the intuitive notion that the energy of an asymptotic signal is concentrated in a finite domain in the general time-frequency plane.

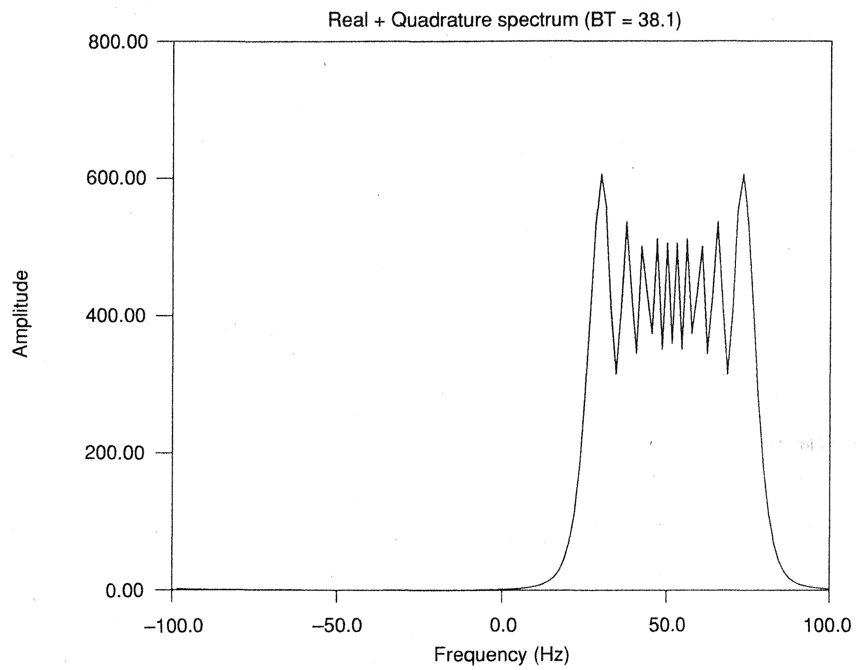

Figure 2.5. The real signal plus quadrature signal spectrum for signal 2 of example 1.

3 The IF and time-frequency distributions

The following section examines the relationship between the IF of a monocomponent signal and its time-frequency representation.

While working on the concepts of the IF and the analytic signal, and borrowing previous results from quantum mechanics, Ville formulated a distribution of a signal in time and frequency referred to as the Wigner-Ville Distribution (WVD) [4]:

$$W(t,f) = \int_{-\infty}^{+\infty} z(t+\tau/2)z^*(t-\tau/2)e^{-j2\pi f\tau}d\tau \tag{30}$$

It was shown in [4] that the first moment of the WVD with respect to frequency yields the IF:

$$f_i(t) = \frac{\displaystyle\int_{-\infty}^{+\infty} fW(t,f)df}{\displaystyle\int_{-\infty}^{+\infty} W(t,f)df} \tag{31}$$

Ideally, one would expect from a time-frequency representation of

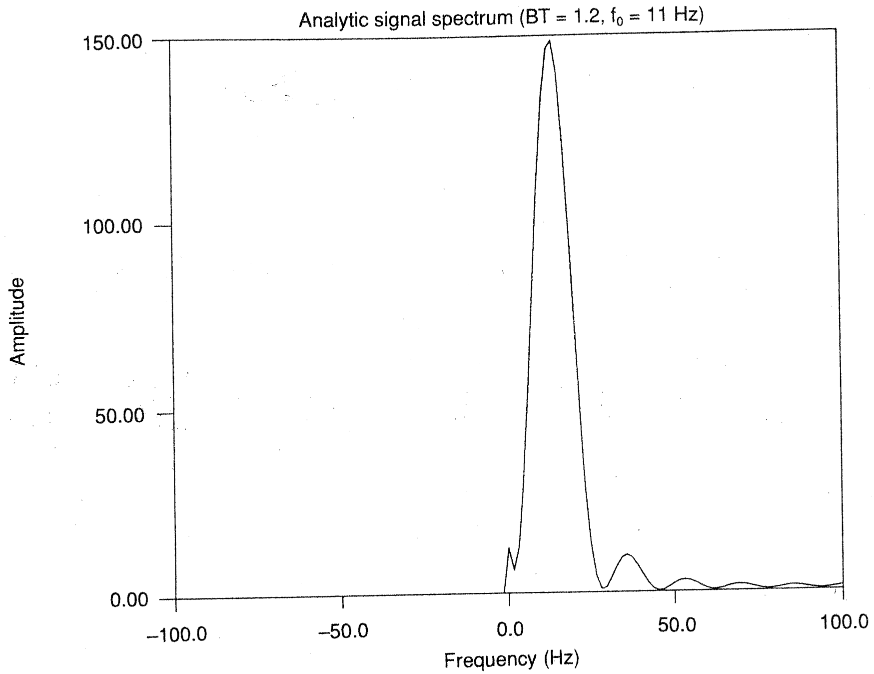

Figure 2.6. The analytic signal spectrum for signal 1 of example 2.

a signal with a large BT product, that the energy would concentrate about the IF, with a spread somehow related to the envelope of the signal. Accordingly, for monocomponent signals of the form of $a(t)e^{j\phi(t)}$ it would be intuitively satisfying to generate a TFD that could be expressed as:

$$\rho(t,f) = A(t,f) \underset{(f)}{*} \delta(f - f_i(t)) \tag{32}$$

where $A(t,f)$ is the time-frequency representation of the amplitude function, $a(t)$. Thus the amplitude and phase would be separable, providing an easily interpretable distribution: the distribution is centred around the varying IF about which the amplitude information is spread in time and frequency.

Unfortunately, no TFD exists that can be expressed in the form given in (32). For example the WVD (for a signal with a large BT product) yields a distribution of the general form [22]:

$$W(t,f) = kA_i\left(\left[\frac{32\pi^2}{f_i''(t)}\right]^{\frac{1}{3}}(f - f_i(t))\right) \tag{33}$$

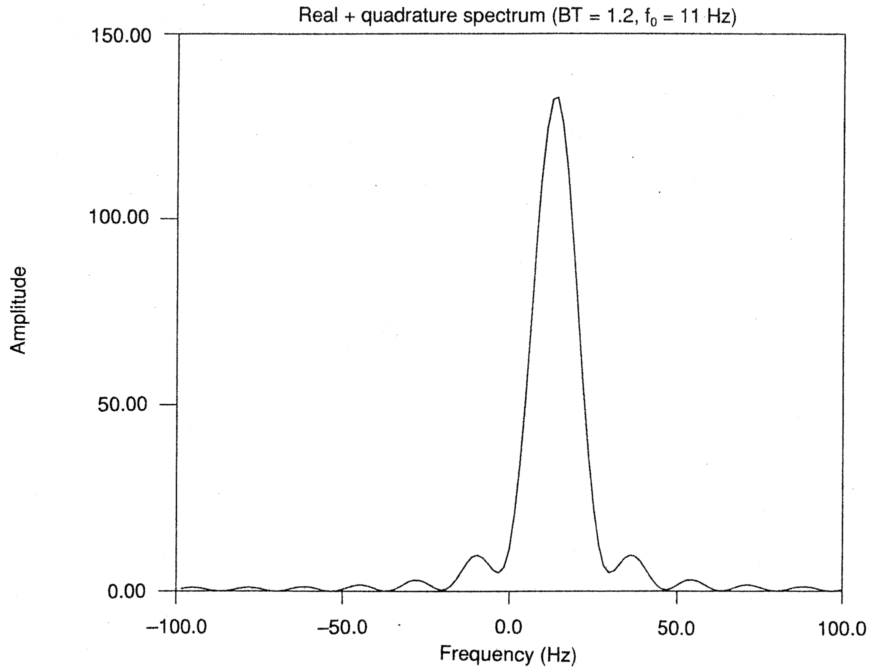

Figure 2.7. The real signal plus quadrature signal spectrum for signal 1 of example 2.

where A_i is the Airy function, and k ensures that $\int_{-\infty}^{+\infty} W(t,f)df = a^2(t)$.

Only for the case where the third order derivative of the phase function is zero (linear FM signals), does the WVD produce a distribution of the form of (32). In other cases, where the phase is more complicated, the WVD is distorted by the Fourier transforms of higher order phase derivative terms.

Many popular TFDs may be calculated by a single generalised formula, derived by Cohen [23]:

$$\begin{aligned}\rho(t,f) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{j2\pi\nu(u-t)}g(\nu,\tau) \\ &\qquad z(u+\tau/2)z^*(u-\tau/2)e^{-j2\pi f\tau}d\nu du d\tau\end{aligned} \tag{34}$$

where $g(v,\tau)$ is the kernel function which defines the particular TFD chosen. The IF will be given by the first frequency moment of any TFD satisfying (34), provided the following conditions hold [15] [23]:

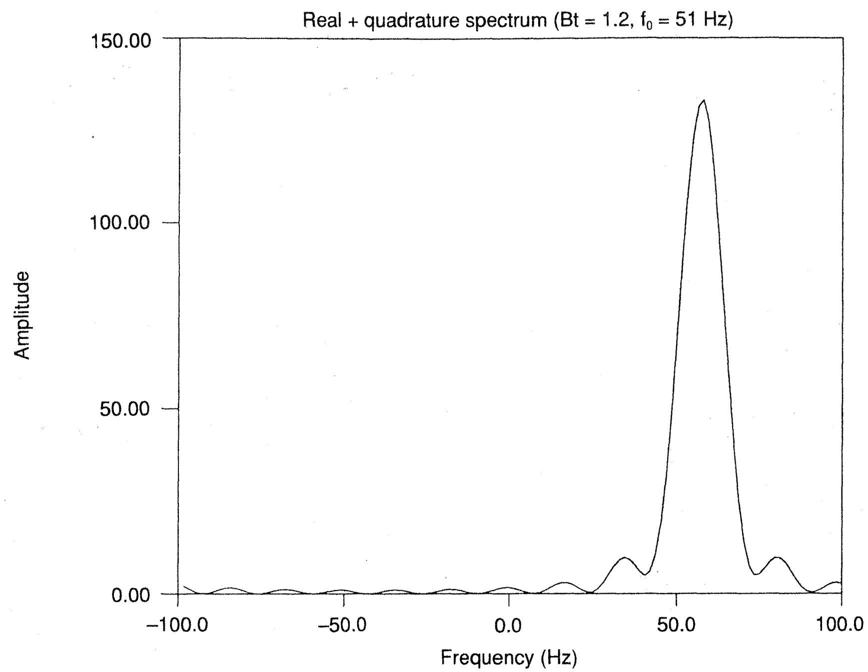

Figure 2.8. The analytic signal spectrum for signal 2 of example 2.

$$g(\nu, 0) = 1 \quad \text{and} \quad \left.\frac{\partial g(\nu, \tau)}{\partial \tau}\right|_{\tau=0} = 0 \tag{35}$$

These conditions are in addition to those that must be satisfied by $g(v, \tau)$ to be a member of Cohen's class [24].

Many distributions yield the IF by correct first moment calculation as discussed above, but this is often computationally expensive and adversely affected by noise. It would be an advantageous feature of a distribution to allow estimation of the IF by peak detection. With signals whose phase functions are quadratic, the WVDs will be of the form of (32) and thus IF estimation can be achieved by peak detection. Such a method has been shown to be optimal for high to modest signal-to-noise ratios (SNRs) [24].

The magnitude squared short-time Fourier transform (STFT) has for a long time been the standard means of analysing non-stationary signals. It may be expressed as [15]:

$$|S_z(t, f)|^2 = \left| \int z(\tau) h(t - \tau) e^{-j2\pi f t} d\tau \right|^2 \tag{36}$$

where $h(t)$ is the analysing window.

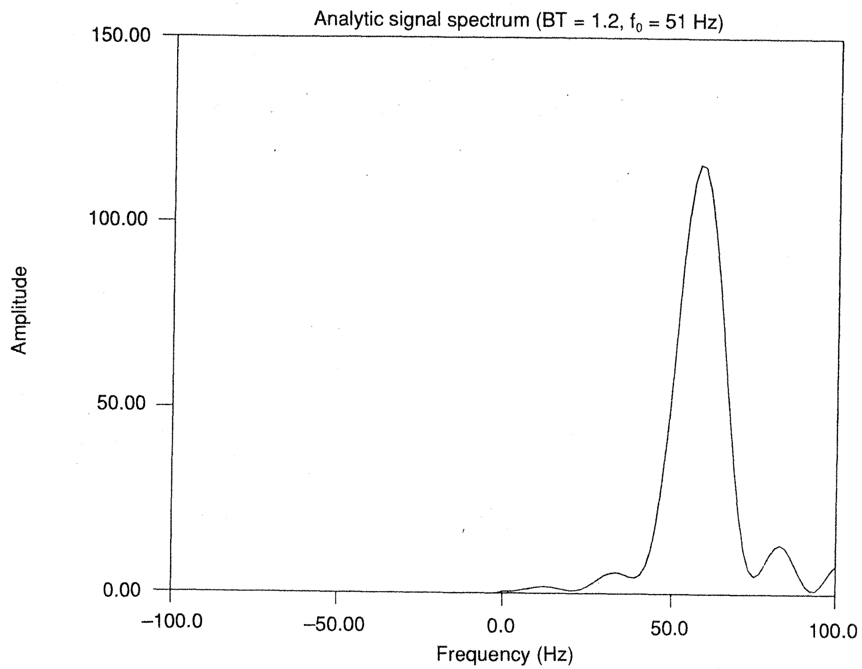

Figure 2.9. The real signal plus quadrature signal spectrum for signal 2 of example 2.

As the STFT is a straightforward extension of the stationary Fourier transform, the representation of a signal and its IF law is dependent on the length of the analysing window. It may be shown that for a signal whose IF law varies by an amount δf_i during the time interval δt, the IF can be tracked (by peak detection) to within an accuracy Δf provided that [18]:

$$\left| \frac{\delta f_i}{\delta t} \right| \leq (\Delta f)^2 \tag{37}$$

where Δf would be the spectral resolution width of the window, $h(t)$. For a signal with a linear IF law, this reduces to a simple expression for the optimum window width [26] [27]:

$$L_h(t) = \left| \frac{df_i(t)}{dt} \right|^{-\frac{1}{2}} \tag{38}$$

This is the largest length window which will ensure that the IF is resolved as an unbiased peak in the STFT. The window must be small enough so that the IF does not vary too much (i.e., the IF is approximately stationary during the analysis time) and large enough to provide the best possible spectral resolution.

3.1 The concept of a multicomponent signal

The signals discussed so far are monocomponent. If the signal is composed of several components, the IF as previously defined does not always have immediate physical significance. Consider the case of a signal:

$$s(t) = a_1(t)\cos\phi_1(t) + a_2(t)\cos\phi_2(t) + \cdots \tag{39}$$

Assuming that each term of the signal $s(t)$ has a large BT product, application of the HT to produce the analytic signal approximately gives:

$$z_m(t) = a_z(t)e^{j\phi_z(t)} = a_1(t)e^{j\phi_1(t)} + a_2(t)e^{j\phi_2(t)} + \cdots \tag{40}$$

The IF of this signal is:

$$f_i(t) = \frac{\sum_{i=1}^{n} a_i(t)^2 f_{i_i}(t) + \sum_{i,j=1[i\neq j]}^{n} \frac{1}{2} q_{i,j}(f_{i_i}(t) + f_{i_j}(t))}{\sum_{i=1}^{n} a_i(t)^2 + \sum_{i,j=1[i\neq j]}^{n} q_{i,j}(t)} \tag{41}$$

where $q_{i,j}(t) = a_i(t)a_j(t)\cos(\phi_i(t) - \phi_j(t))$ and the derivatives of the amplitude terms are assumed negligible.

Little or no meaning can be attached to this IF value, which is oscillatory and highly non-linear, and it is in such situations that the concept of a multicomponent signal (and indeed a monocomponent signal) becomes attractive. Thus, this tentative intuitive definition of a multicomponent signal is proposed:

Definition

A finite energy asymptotic signal, $s(t)$, is referred to as a multicomponent signal if there exists a finite number, N, of monocomponent signals, $s_i(t)$ $(i = 1, N)$, such that the relation $s(t) = \sum_{i=1}^{N} s_i(t)$ holds for all values of t for which $s(t)$ is defined, and this decomposition is meaningful.

Naturally, this definition must be applied with care as there can be no unique decomposition into individual monocomponent signals without a priori knowledge of the signal.

A consideration of mono and multicomponent signals, in conjunction with TFDs, leads to further interesting results and observations. The concept of a multicomponent signal is both a local and a global phenomenon. As figure 2.10 illustrates, the signal is multicomponent everywhere except where the two signals cross, and at that moment the signal is truly monocomponent. The concept of a multicomponent signal does have a global significance as the real signal must be known

for all time to construct the analytic signal and the distribution. Additionally, the individual signals in a multicomponent signal do not have unique time-frequency laws unless the signal is always multicomponent (i.e., the signals' laws do not cross). As an example consider the signals displayed in figure 2.10 — two possible decompositions into monocomponent signals are shown.

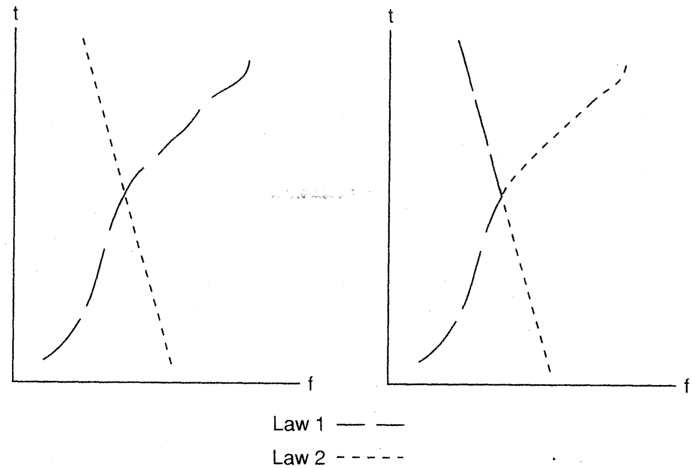

Figure 2.10. Alternate ways of separating two frequency laws (decomposition of multicomponent signals is not unique unless the frequency laws of the components do not cross at any point).

4 IF estimation via TFDs

A number of approaches may be adopted for estimation of the IF of a non-stationary process. We present here the methods which use TFDs directly.

4.1 IF estimation using the first moment of a TFD

An approach to the estimation of IF, which uses moments of TFDs, and in particular, the WVD, was proposed in [30]. Taking the first moment of the general TFD expressed in equation (34) yields [15]:

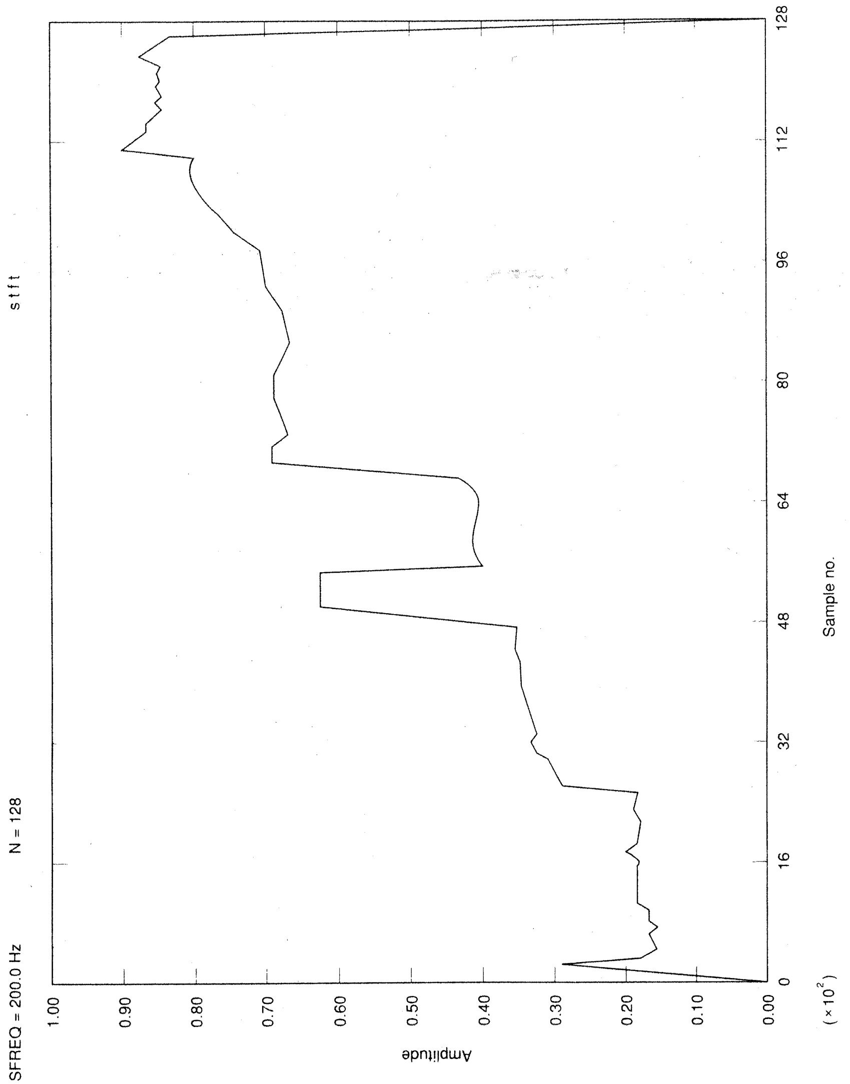

Figure 2.11. The IF estimate of the reference signal in OdB noise, the estimate being obtained from the peak of the STFT.

$$
\begin{aligned}
m_\rho^1(t) &= \frac{\displaystyle\int_{-\infty}^{+\infty} f\rho(t,f)df}{\displaystyle\int_{-\infty}^{+\infty} \rho(t,f)df} \\
&= \frac{1}{2\pi}\frac{\displaystyle\int_{-\infty}^{+\infty} G(u-t,0)\mathrm{Im}(dz/dt)z^*(u)du}{\displaystyle\int_{-\infty}^{+\infty} G(u-t,0)|z(u)|^2du}
\end{aligned} \tag{42}
$$

where $G(t,\tau)$ is the Fourier transform of the kernel function, $g(v,\tau)$ of (34).

This equation shows that the first moment of a TFD is, in general, a smeared version of the IF. If the signal magnitude is constant, then:

$$m_\rho^1(t) = f_i(t) \underset{(t)}{*} G(t,0) \tag{43}$$

Thus in general the operation of estimating the IF by use of the first frequency moment of a TFD results in a smoothed version of the IF:

$$f_i^s(t) = f_i(t) * h(t) \tag{44}$$

where $h(t) = G(t,0)$.

For the case of the WVD, $G(t,0) = \delta(t)$ and there is no smoothing, whereas for an STFT, $h(t)$ becomes the original analysing window [31] and the result is a smoothed estimate of the IF, which when noise is present has a low variance but is characteristically biased [15] [39]. An examination of the STFT and its representation of IF, including results for choosing optimal windows, are given in chapter 8 of this book by Harris.

Due to the high computational complexity, IF estimation utilizing the first moment of a TFD is not always a preferred method. The particular advantage of this technique, however, lies in the preprocessing that may be performed in the time-frequency plane (see section 5.1) — noise effects may be reduced as well as possibly allowing IF estimation of multicomponent signals.

4.2 Peak detection from TFDs for IF estimation

Optimal estimation of the frequency of a sinusoidal process in Gaussian noise is achieved by peak detection from the magnitude squared STFT [33]. Where the signal is a non-stationary process, it would seem reasonable to employ a similar method to estimate

the instantaneous frequency (IF). The short-time Fourier transform (STFT), however, has poor energy concentration properties for rapidly time-varying signals [15]. IF estimation using the peak of the WVD may be used. The performance of this estimator, however, significantly degrades at low SNR [34]. The cross WVD (XWVD), which achieves high energy concentration as well as excellent noise preformance, was proposed for obtaining an IF estimate [18]. The XWVD between a reference signal $z_s(t)$ and an observed signal $z_r(t)$ is:

$$W_{z_s,z_r} = \int_{-\infty}^{\infty} z_s(t+\tau/2) z_r(t-\tau/2) e^{-j2\pi f\tau} d\tau \qquad (45)$$

The following recursive IF estimation procedure is then:

1 Obtain an initial estimate of the IF [15] and form a unit amplitude signal that has this IF estimate as its frequency law;

2 Using this as reference form a XWVD and extract the peak (of the squared XWVD) as the new IF estimate; and

3 Repeat the procedure from step 1.

Each time a new XWVD is estimated, the signal energy is more concentrated, so that there is a greater probability that its peak will be correctly estimated from a background of noise. For chirps at high SNR, the performance of the XWVD scheme will approach the performance of the STFT for estimating a corresponding sinusoidal frequency. As the SNR increases, the estimate will still exhibit very low variance compared with spectrogram estimates, provided the SNR and the spectral variation obey certain criteria. An error analysis and convergence performance are provided in [18]. As an example consider a linear FM signal in 0 dB white noise. The initial estimate of the IF (using peak detection from an STFT) is shown in figure 2.11. The estimate after one iteration of the above-described method is displayed in figure 2.12 (The true IF is shown in figure 2.13). It may be seen that the updated estimate is a significant improvement over the original. The algorithm typically converges in a few iterations.

5 Instantaneous frequency estimation for discrete time signals

This section provides a brief introduction to the problem of IF estimation for discrete time signals. A comprehensive study can be found in [39] and [42].

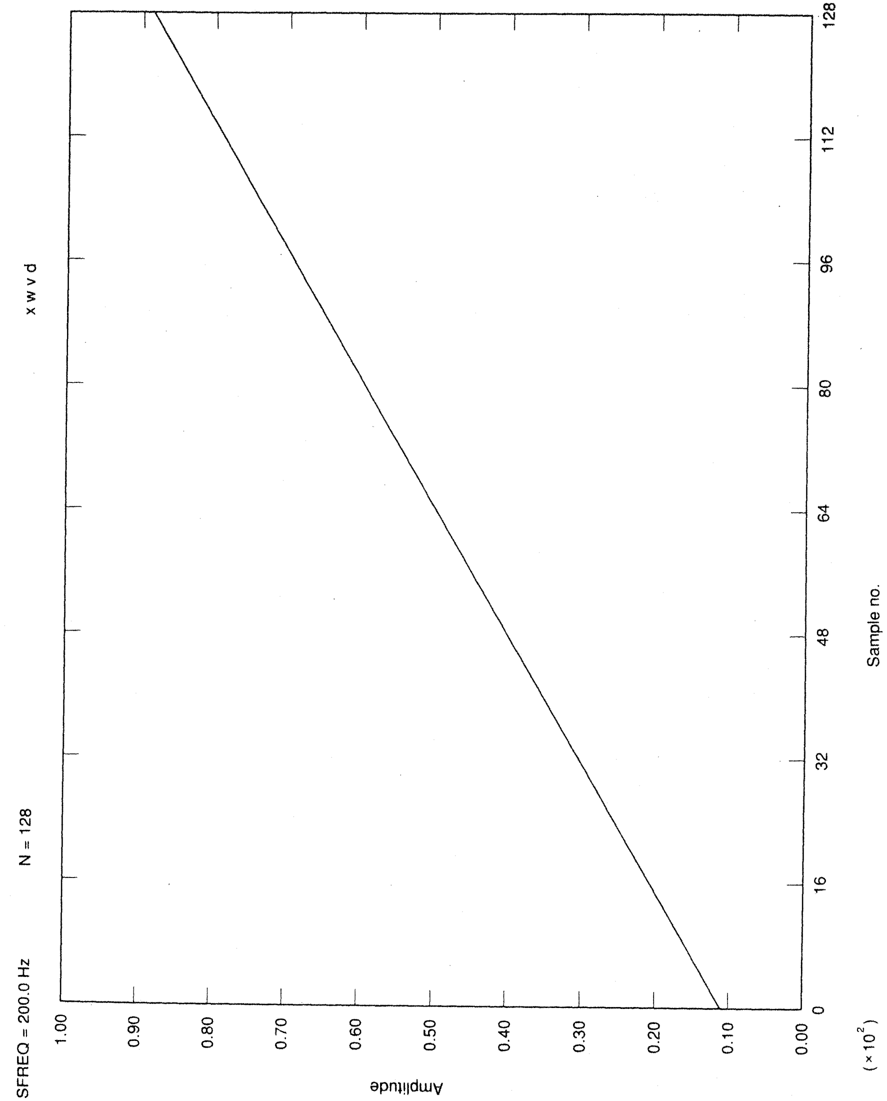

Figure 2.12. The IF estimate of the reference signal obtained from the XWVD scheme after one iteration.

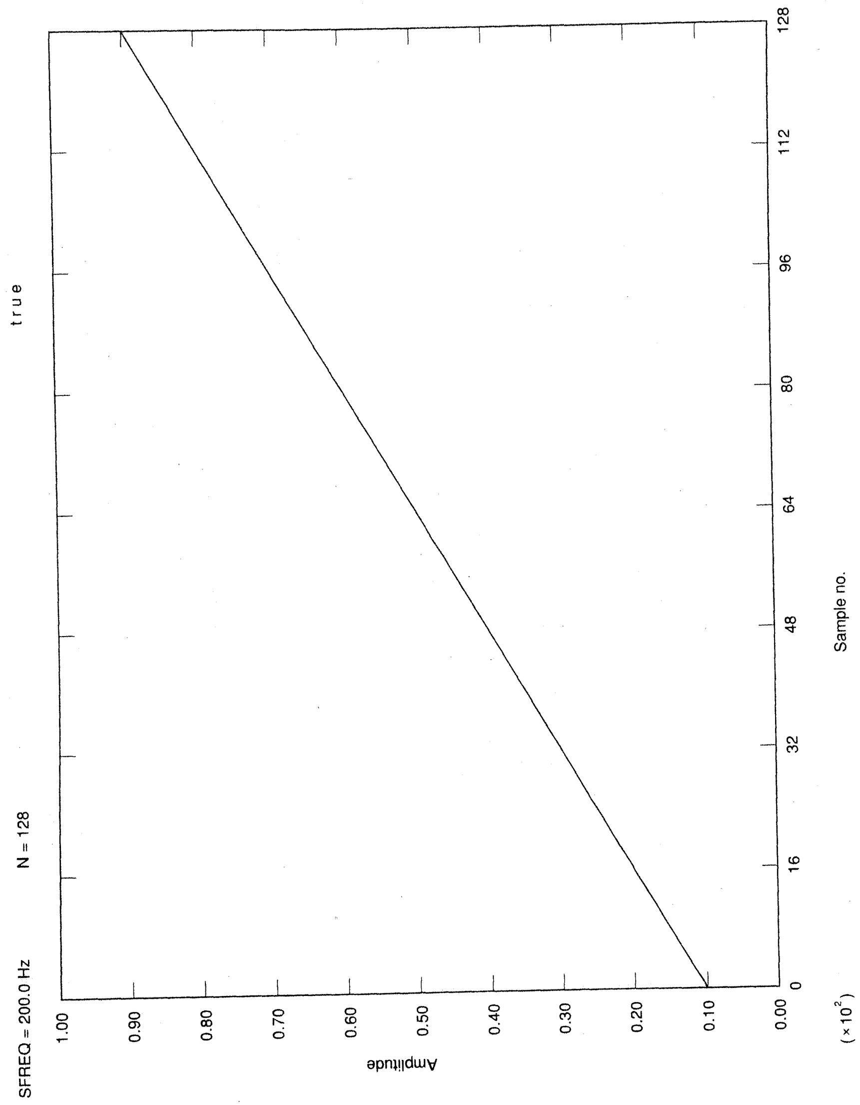

Figure 2.13. The true IF of the reference signal.

5.1 Discrete IF measurements based on finite differencing

The expression for the IF of a continuous time complex signal, $z(t)$, may be re-written as:

$$f_i(t) = \lim_{\delta t \to 0} \frac{1}{4\pi\delta t}(\phi(t+\delta t) - \phi(t-\delta t)) \tag{46}$$

This may be translated directly to the discrete time case which provides an estimate of discrete instantaneous frequency (DIF) — where for example a central finite difference operation replaces the differentiation [19, part II]:

$$\hat{f}_i(n) = \frac{f_s}{4\pi}(\phi(n+1) - \phi(n-1)) \tag{47}$$

It is assumed here that the discrete phase, $\phi(n)$, has been unwrapped [39]. The discrete time analytic signal, $z(n)$, associated with the real discrete time signal, $s(n)$, is given by:

$$z(n) = s(n) + jH[s(n)] \tag{48}$$

where $H[\ ]$ is the discrete time Hilbert transform [20] which may be defined by:

$$H[s(n)] = \sum_{m=-\infty}^{+\infty} \frac{2s(n-m)}{m\pi}, \quad \text{(m odd)} \tag{49}$$

Other estimates based on finite differencing have been proposed [21]. For a discrete time signal of the following form:

$$x(n) = A\cos\phi(n) \tag{50}$$

Griffith estimates the DIF as:

$$\hat{f}_i(n) = \frac{f_s}{2\pi}(\phi(n) - \phi(n-1)) \tag{51}$$

which is convenient for signals with modulation laws expressable as:

$$\hat{f}_i(n) = f_0 + m(n) \tag{52}$$

Claasen and Mecklenbräuker indicate in [19, part II] that the central finite difference DIF of (47) is more appropriate than the backward difference estimator (equation 51) or the forward difference $(\phi(n+1) - \phi(n))$ since the latter are asymmetrical.

The central finite difference does not always produce unbiased estimates of the DIF. For IF laws with a non-linear variation, the estimate will be biased. An unbiased estimate for the DIF which may be expressed as a generalised finite difference estimator is discussed in [39] and [42].

5.2 Zero-crossing measures

A traditional and very simple method for estimating the DIF is the zero-crossing detector, which detects the discrete time distance between adjacent zero crossings, and then makes a frequency estimate assuming a stationary sinusoid:

$$f = \frac{1}{2T_z} \tag{53}$$

where T_z is the zero crossing interval.

This estimator is straightforward to implement, is naturally suboptimal and will produce biased estimates for non-stationary IF laws. Various forms of averaging may be used to reduce the estimator's variance [41].

5.3 Short-time estimates

A number of estimation techniques have been proposed based on the assumption that the DIF is stationary over the local analysis time. As a straightforward extension of the periodogram, the magnitude squared STFT provides an estimate of the DIF by peak detection. This estimate becomes maximum likelihood in high SNR for a linear DIF, providing the analysis window has been appropriately selected [39].

The resolution of such estimates may be improved by the use of parametric techniques. Griffiths [21] used a least mean square (LMS) linear prediction based spectral estimate. The prediction filter coefficients are recursively updated to reduce the computations required and preserve the local character of the estimates. The method is unable to resolve fast changing DIFs, but as a result it has good noise rejection properties for slowly varying laws. Other techniques, based on recursive least squares (RLS) approaches to the linear prediction problem, and time-varying autoregressive models, have been suggested [28].

5.4 Polynomial phase modelling

A methodology for DIF estimators has been proposed based on polynomial phase modelling in [39] and [42]. It is assumed that the phase of the discrete signal may be modelled as a polynomial of order p:

$$\phi(n) = a_0 + a_1 n + a_2 n^2 + \cdots + a_p n^p \tag{54}$$

The polynomial coefficients may be estimated in a number of ways. If the phase is unwrapped, linear least squared techniques may be used

to solve for the polynomial coefficients (for this case it is assumed the SNR is high and the noise is additive zero-mean Gaussian). By generalising the result for estimating stationary tone parameters it is possible to derive a maximum likelihood estimator for the coefficients. Adaptive techniques may be incorporated into this estimator, and it may be implemented efficiently provided the polynomial order is small. Explicit details of these phase estimators as well as simulated examples may be found in [39] and [42].

6 Applications

6.1 Automatic time-varying filtering

The use of TFDs for non-stationary signal analysis has the advantage that a priori or a posteriori processing in the time-frequency plane can be employed to give enhanced estimates. This ability to perform time-frequency filtering can be applied to tasks such as signal enhancement and IF estimation. To perform this time-frequency filtering, a time-varying system transfer function, $H(t, f)$, must be specified in accordance with the application. The WVD, being an entirely real transform, is well suited to time-varying filtering, particularly for monocomponent signals. The time-varying filtering operation is given by:

$$y(t) = W^{-1}[W(t, f) \cdot H(t, f)] \tag{55}$$

where $W(t, f)$ is the WVD to be processed, and $W^{-1}[\ ]$ represents the synthesis operation. This operation determines the time domain analytic signal whose WVD is closest to the operand in a least mean squares sense [36]. If the operand is a valid WVD, the synthesis operator is equivalent to WVD inversion. For an in depth discussion, the reader is referred to chapter 17 on signal synthesis.

For the analysis of a monocomponent signal, the time-varying filter would ideally capture all the signal energy and reject all the noise. In practice, one must settle for a tradeoff in which the significant signal energy is captured, while most of the noise is eliminated. A useful approach is to window about the IF, since it has been previously explained that, for monocomponent signals, the signal energy tends to be concentrated there.

The various techniques proposed earlier, then, may be used for IF estimation as a first step in the design of the automatic filter. The bandwidth of the time-varying filter may be selected according to the application, as can the window type. For example, consider the problem of estimating the IF of a linear FM signal distorted by 3 dB white Gaussian noise. Figure 2.14 shows the IF law estimated using

the direct definition, while figure 2.15 shows the IF law estimated by using the first frequency moment of the WVD after a time-frequency window has been applied. In a practical situation one could start by obtaining a rough estimate of the IF law, then window about this region of time-frequency concentration, and finally obtain the improved IF estimate. One could also use signal synthesis to obtain a time domain signal from the windowed WVD, and calculate the IF directly from this filtered signal.

6.2 Some specific areas of application

In a typical seismic situation, a vibroseis signal is emitted into the earth. The different arrival times of reflected waveforms are used to reveal information as to the geological structure of the region. The reflected returns can also be processed to yield information as to the amount of dispersion present in the signal. This can be determined simply by comparing the IF of the received signal with that of the source [38] (see chapter 20 of this book on seismic application).

In the fields of sonar and radar, the classical problem of direction-of-arrival estimation may be reduced to a problem of estimation of multiple IFs [40]. In fact, the estimation of IF is related to almost any problem that involves the analysis or tracking of a narrowband time-varying signal. This includes such natural phenomena as the echolocation of bats and sounds of cetacea, and more specialised applications such as the analysis of ECGs and helicopter blade rotations as presented in chapter 18 of this book by D. Forrester. Although the IF may not be specifically emphasised, it is very often the quantity of interest.

7 Summary

Instantaneous frequency estimation is an important problem in signal processing. It raises fundamental questions in signal representation, some of which have been addressed in this chapter.

The original developments leading to the concept of 'instantaneous frequency' have been reviewed, and the relation of the IF to TFDs has been stressed. The concepts of monocomponent and multicomponent signals have been discussed. Techniques of IF estimation have been presented and some applications of IF estimation have been described.

Acknowledgements

The authors acknowledge the general support of the Defence Science and Technology Organisation (DSTO) Australia, and the Australian Research Council (ARC).

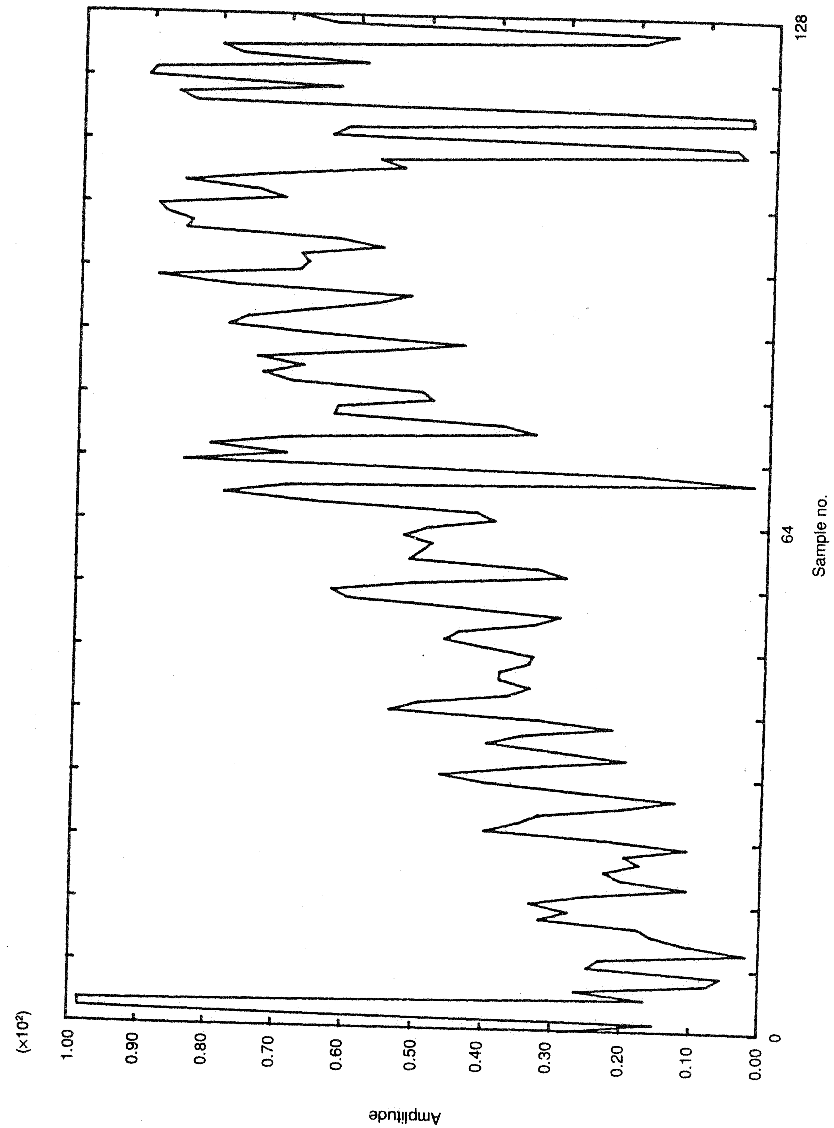

Figure 2.14. IF estimate obtained from the phase of the analytic signal.

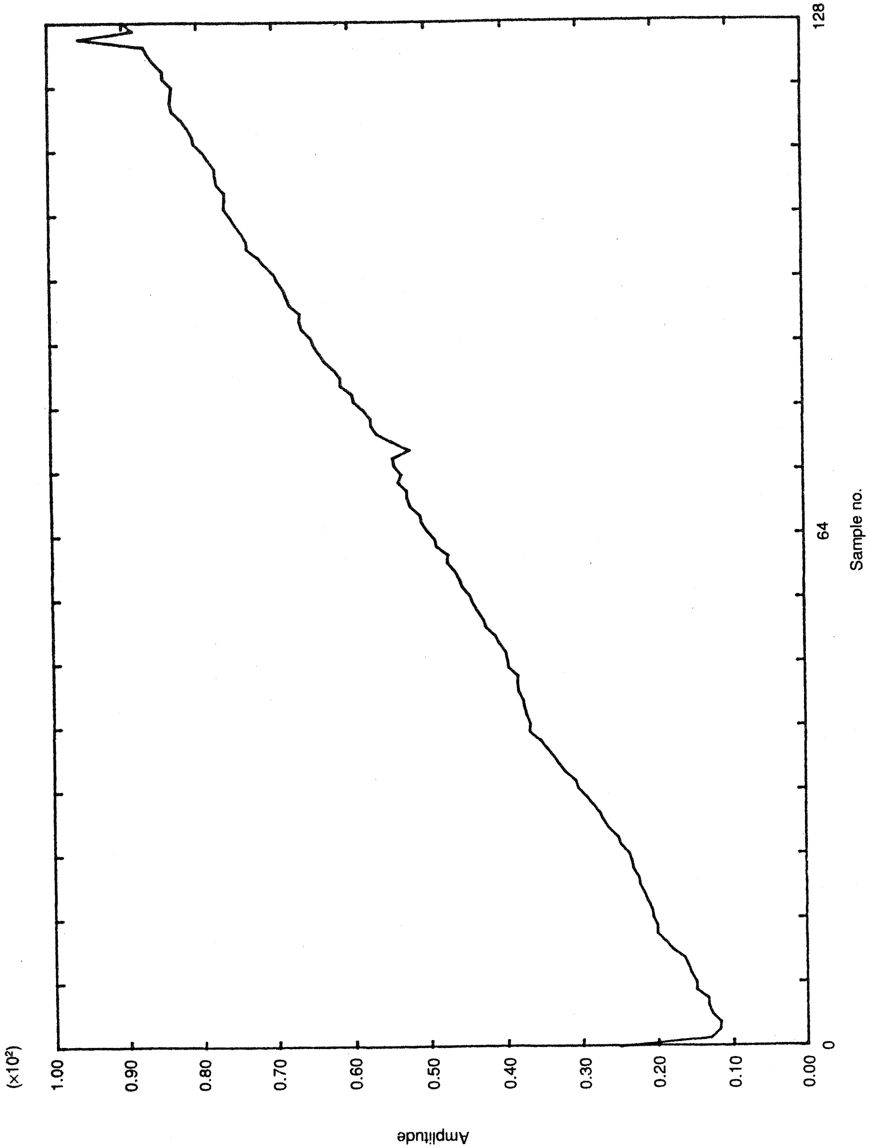

Figure 2.15. IF estimate obtained from the filtered WVD's first moment (the filtering has reduced the effects of outlying noise).

References

[1] J. Carson and T. Fry, 'Variable Frequency Electric Circuit Theory with Application to the Theory of Frequency Modulation', *Bell System Tech. J.,* 16, pp. 513-540, 1937.

[2] B. Van der Pol, 'The Fundamental Principles of Frequency Modulation', *Proc. of IEE*, 93 (III), pp. 153-158, 1946.

[3] D. Gabor, 'Theory of Communication', *Proc. IEE*, 93 (III), pp. 429-457, 1946.

[4] J. Ville, 'Theorie et Application de la Notion de Signal Analytic', *Cables et Transmissions*, 2A, pp. 61-74, 1948.

[5] B. Boashash, 'Note on the Use of the Wigner Distribution for Time-Frequency Signal Analysis', *IEEE Trans. on ASSP*, vol.36, pp. 1518-1521, Sept., 1988.

[6] J. Shekel, 'Instantaneous Frequency', *Proc. of IRE*, 41, pp. 548, 1953.

[7] L. Mandel, 'Interpretation of Instantaneous Frequency', *Amer. J. Phys.*, 42, pp. 840-846, 1974.

[8] A. Rihaczek, 'Signal Energy Distribution in Time and Frequency', *IEEE Trans. on Information Theory*, IT-14, pp. 369-374, 1968. part III, John Wiley, 1969.

[9] M.H. Ackroyd, 'Instantaneous and Time-Varying Spectra — An Introduction', *Journal of the Radio and Electronic Engineer*, vol.39, no.3, 1970.

[10] M.S. Gupta, 'Definition of Instantaneous Frequency and Frequency Measurability', *American Journal of Physics*, vol.43, no.12, December 1975.

[11] K. Kodera, R. Gendrin and C. Villedary, 'Analysis of Time-Varying Signals with Small BT Products', *IEEE Trans. ASSP,* 26, pp. 64-76, 1978.

[12] E. Bedrosian, 'A Product Theorem for Hilbert Transforms', *Proc. IEEE*, 51, pp. 686-689, 1963.

[13] A. Rihaczek, 'Hilbert Transforms and the Complex Representation of Signals', *Proc. IEEE*, 54, pp. 434-435, 1966.

[14] A. Nuttall, 'On the Quadrature Approximation to the Hilbert Transform of Modulated Signals', *Proc. IEEE*, 54, pp. 1458, 1966.

[15] B. Boashash, 'Time-Frequency Signal Analysis', *Advances in Spectrum Estimation*, edited by S. Haykin, Prentice-Hall, 1990, pp. 418-517.

[16] H. Landau, D. Slepian and H. Pollack, 'Prolate Spheroidal Wave Functions — Fourier Analysis and Uncertainty', *Bell System Technical Journal,* vol.40, no.1, pp. 43-84, 1961.

[17] D. Vakman, 'Do we Know What are the Instantaneous Frequency and Instantaneous Amplitude of a Signal', *Radio Eng. and Electron, Phys.*, 21(6), pp. 95-100, 1976.

[18] B. Boashash, G. Jones and P. O'Shea, 'Instantaneous Frequency of Signals: Concepts, Estimation Techniques and Applications', *Proc. of SPIE Conf.*, in 'Advanced Algorithms and Architectures for Signal Processing', San Diego, 1989 (See also, G. Jones, PhD Thesis, Queensland University of Technology).

[19] T.A.C.M. Claasen and W.F.G. Mecklenbräuker, 'The Wigner Distribution — Parts I and II', *Philips Research Journal*, vol.35, pp. 217-250 and pp. 276-300, 1980.

[20] A.V. Oppenheim and R.W. Schaffer, *Discrete-Time Signal Processing*, Prentice-Hall, NJ, 1989.

[21] L. Griffiths, 'Rapid Measurement of Digital Instantaneous Frequency', *IEEE Trans. on ASSP*, vol.23, no.2, 1975.

[22] M.V. Berry, 'Semi-Classical Mechanics in Phase Space: A Study of Wigner's Function', *Phil. Trans. Roy. Soc.*, A.287, pp. 237-271, 1977.

[23] L. Cohen, 'Generalised Phase-Space Distributions', *J. Math. Phys.*, 7, pp. 781-786, 1966.

[24] L. Cohen, 'Time-Frequency Distributions', *Proc. IEEE*, V77, no.7, pp. 941-981, 1989.

[25] F.J. Taylor and P. Rao, 'Estimation of Instantaneous Frequency Using the Discrete Wigner Distribution', *Elec. Letters*, vol.40, no.4, pp. 246-248, 1990.

[26] L.R.O. Storey, 'An Investigation of Whistling Atmospherics', *Trans. Roy. Soc.*, A246, pp. 113-141, 1953.

[27] N.F. Barber, F. Ursell, 'The Response of a Resonant System to a Gliding Tone', *Phil. Mag.*, vol.39, pp. 345-361, 1948.

[28] F.J. Harris and H. Abu Salem, 'Performance Comparison of Wigner-Ville Based Techniques to Standard FM Discrimination for Estimating Instantaneous Frequency of a Rapidly Slewing FM Sinusoid in the Presence of Noise', *Proceedings of SPIE Conference, Advanced Algorithms and Architectures for Signal Processing III*, vol.975, pp. 232-244, August, 1988.

[29] D. Hush, N. Ahmed and R. David, 'Instantaneous Frequency Estimation Using Adaptive Linear Predictor Weights', *IEEE Trans. on Aerospace and Electronic Systems*, vol. AES-22, no.4, July 1986.

[30] L. White and B. Boashash, 'Estimating the Instantaneous Frequency of a Gaussian Random Process Using the Wigner-Ville Distribution', *IEEE Transactions on ASSP*, vol.36, no.3, pp. 417-420, March, 1988.

[31] L. Cohen and C.Lee, 'Instantaneous Frequency, its Standard Deviation and Multicomponent Signals', *Proc. of SPIE*, Advanced Algorithms and Architectures for Signal Processing III, vol.975, pp. 186-208, 1988.

[32] K.M. Wong and Q. Jin, 'Estimation of the Time-Varying Frequency of a Signal: The Cramer-Rao Bound and the Application of Wigner Distribution', *IEEE Trans. ASSP*, vol.31, no.3, pp. 519-536, 1990.

[33] S. Kay, *Modern Spectral Estimation*, Prentice-Hall, 1987.

[34] B. Kumar and C. Carroll, 'Performance of Wigner Distribution Function Based Detection Methods', *Opt. Eng.*, 23, 6, pp. 732-737, Nov. 1984.

[35] G. Jones and B. Boashash, 'Instantaneous Frequency and the Analysis of Multicomponent Signals', *Proc. ISSPA 90*, Gold Coast, Australia, 1990, pp. 833-837.

[36] G. Boudreaux-Bartels and T. Parks, 'Time-Varying Filtering and Signal Estimation Using Wigner Distribution Synthesis Techniques', *IEEE Trans. on ASSP*, vol. ASSP-34, no.3, pp. 442-451, 1985.

[37] B. Boashash and L. White, 'Instantaneous Frequency Estimation and Automatic Time-Varying Filtering', *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing,* 90, Albuquerque, New Mexico, April 1990, pp. 1221-1224.

[38] B. Bouachache, 'Wigner Analysis of Time Varying Signals and Application to Seismic Prospecting', EUSIPCO 83, (Erlangen, Germany), North Holland edition, Amsterdam, pp. 703-705, 1983.

[39] B. Boashash, P. O'Shea and M. Arnold, 'Algorithms for Instantaneous Frequency Estimation', *Proc. of SPIE Conf., Advanced Algorithms and Architectures for Signal Processing*, vol.1348, San Diego, 1990.

[40] A. Swindlehurst and T. Kailath, 'Near-Field Source Parameter Estimation Using a Spatial Wigner Approach', *Proc. of SPIE*, Advanced Algorithms and Architectures for Signal Processing III, vol.975, 1988.

[41] L. Rabiner and R. Schafer, 'Digital Processing of Speech Signals', Prentice-Hall, 1978.

[42] B. Boashash, 'Interpreting and Estimating the Instantaneous Frequency of a Signal-Part 1: Fundamentals Part 2: Algorithms', Preprint, *Proceedings of IEEE*, Jan. 1991.